dlp_hw3 实验报告 翻译部分

钟珞妍 PB20061329

1. Abstract:

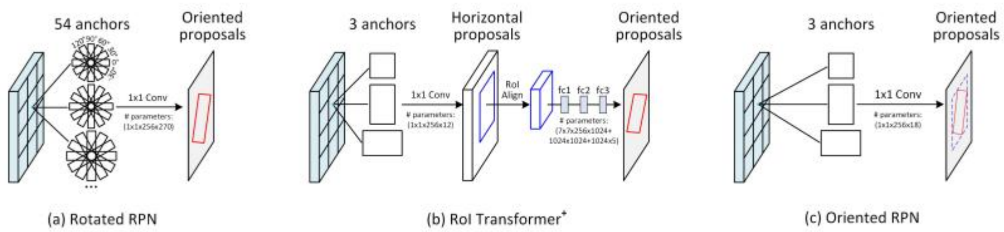
Current state-of-the-art two-stage detectors generate oriented proposals through time-consuming schemes.
当前 SOTA 两阶段检测器生成有向候选框比较费时。
This diminishes the detectors' speed, thereby becoming the computational bottleneck in advanced oriented object detection systems.
这制约了目标检测的速度，从而成为了有向目标检测领域中的计算瓶颈。
This work proposes an effective and simple oriented object detection framework, termed Oriented R-CNN, which is a general two-stage oriented detector with promising accuracy and efficiency.
本文提出了一种有效且简单的有向目标检测方法，称为 Oriented R-CNN，它是一个具有高精度与高效率的两阶段有向目标检测器。
To be specific, in the first stage, we propose an oriented Region Proposal Network (orientedRPN) that directly generates high-quality oriented proposals in a nearly cost-free manner.
具体来说，在第一阶段，我们提出了 orientedRPN，它能几乎不费时地产生高质量的有向候选框。
The second stage is oriented R-CNN head for refining oriented Regions of Interest (orientedRoIs) and recognizing them.
在第二阶段，是 orientedRoIs，用于分类和回归。
Without tricks, oriented R-CNN with ResNet50 achieves state-of-the-art detection accuracy on two commonly-used datasets for oriented object detection including DOTA (75.87% mAP) and HRSC2016 (96.50% mAP), while having a speed of 15.1 FPS with the image size of 1024×1024 on a single RTX2080Ti.
在没有附加功能的情况下，以 ResNet50 作为 backbone，oriented R-CNN 在 DOTA 和 HRSC2016 上都取得了 SOTA 的精确度，DOTA(75.87% mAP)，HRSC2016(96.50%mAP)，并且输入大小为 1024x1024 的图像，在单张 RTX2080Ti 上训练能够达到 15.1FPS 的速度。
We hope our work could inspire rethinking the design of oriented detectors and serve as a baseline for oriented object detection.
我们希望本文能启发大家重新思考旋转检测器的设计并且作为旋转目标检测领域的一个 baseline。
Code is available at https://github.com/jbwang1997/OBBDetection.

代码开源在 https://github.com/jbwang1997/OBBDetection 。

2. Introduction:

Most existing state-of-the-art oriented object detection methods depend on proposal-driven frameworks, like Fast/Faster R-CNN. They involve two key steps: (i) generating oriented proposals and (ii) refining proposals and classifying them into

different categories. Nevertheless, the step of producing oriented proposals is computationally expensive
大多现有的 SOTA 旋转目标检测算法是基于候选框的，例如 Fast/Faster R-CNN，它们包括两步：(i)产生有向候选框 (ii)微调候选框并进行分类，然而，产生候选框的过程十分耗时。
One of the early methods of generating oriented proposals is rotated Region Proposal Network (rotated RPN for short), which places 54 anchors with different angles, scales and aspect rations (3 scales×3 ratios×6 angles) on each location, as shown in Figure 1(a). The introduction of rotated anchors improves the recall and demonstrates good performance when the oriented objects distribute sparsely. However, abundant anchors cause massive computation and memory footprint. To address this issue, RoI Transformer learns oriented proposals from horizontal RoIs by complex process, which involves RPN, RoI Alignment and regression (see Figure 1(b)). The RoI Transformer scheme provides promising oriented proposals and drastically reduces the number of rotated anchors, but also brings about expensive computation cost. Now, how to design an elegant and efficient solution to generate oriented proposals is the key to breaking the computational bottleneck in state-of-the-art oriented detectors.
早期的一种产生有向候选框的方式是 rotated RPN，这种方法在每个位置会放置 3scales x 3ratios x 6angles 个不同的 anchor，如图 1(a)所示。这种 rotated anchor 能够提高召回率并且在旋转目标较为稀疏的时候有很好的表现。但是，数量众多的 anchors 带来了极大的计算量和内存消耗。为了解决这个问题，RoI Transformer 通过复杂的过程从水平 RoI 中学习生成有向候选框，这个过程包括 RPN，RoI Alignment 和 regression，如图 1(b)所示。RoI Transformer 方案确实能够提供好的有向候选框并极大地减少了 rotated anchors 的数量，但是仍然具有很大的计算成本。目前，如何设计一个优雅且高效的方案来产生有向候选框是突破当前 SOTA 旋转检测计算瓶颈的关键。



To push the envelope further: we investigate why the efficiency of region proposal-based oriented detectors has trailed thus far. Our observation is that the main obstacle impeding the speed of proposal-driven detectors is from the stage of proposal generation. A natural and intuitive question to ask is: can we design a general and simple oriented region proposal network (oriented RPN for short) to generate high-quality oriented proposals directly? Motivated by this question, this paper presents a simple two-stage oriented object detection framework, called oriented R-CNN, which obtains state-of-the-art detection accuracy, while keeping competitive efficiency in comparison with one-stage oriented detectors.
为了进一步推进这个问题的解决，我们调查了有向目标检测到目前为止一直效率较低的原因。我们观察到阻碍基于候选框的检测器计算速度的原因在于产生候选

框的阶段。自然的，我们会问：能否设计一个简单的 oriented RPN 来直接产生高质量的有向候选框？基于这个问题，本文提出了一种简单的两阶段有向目标检测器的框架，称作 oriented R-CNN，它具有 SOTA 的精确度并且可匹敌一阶段目标检测器的效率。

To be specific, in the first stage of oriented R-CNN, a conceptually simple oriented RPN is presented (see Figure1(c)). Our oriented RPN is a kind of light-weight fully convolutional network, which has extremely fewer parameters than rotated RPN and RoI transformer^{+}, thus reducing the risk of overfitting. We realize it by changing the number of output parameters of RPN regression branch from four to six. There is no such thing as a free lunch. The design of oriented RPN benefits from our proposed representation scheme of oriented objects, named midpoint offset representation. For arbitrary-oriented objects in images, we utilize six parameters to represent them. The midpoint offset representation inherits horizontal regression mechanism, as well as provides bounded constraints for predicting oriented proposals. The second stage of oriented R-CNN is oriented R-CNN detection head: extracting the features of each oriented proposal by rotated RoI alignment and performing classification and regression.

具体来说，在 oriented R-CNN 的第一阶段，是一个简单的 oriented RPN，如图 1(c)所示。这个 oriented R-CNN 是一个轻量化的全卷积网络，比 rotated RPN 和 RoI transformer 所需的参数都少的多，从而减小了过拟合的风险。我们通过将 RPN 回归分支的输出参数从 4 个改为 6 个来实现，天下没有免费的午餐。这种 oriented RPN 的设计得益于我们提出的有向目标的表示方法：中点偏移表示法。我们使用 6 个参数来表示图像中任意角度的物体。中点偏移表示法借鉴了水平回归机制，并为有向候选框的预测提供了有界约束。Oriented R-CNN 的第二阶段是一个检测头，通过 rotated RoI alignment 和分类、回归来实现对每个有向候选框的特征提取。

Without bells and whistles, we evaluate our oriented R-CNN on two popular benchmarks for oriented object detection, namely DOTA and HRSC2016. Our method with ResNet-50-FPN surpasses the accuracy of all existing state-of-the-art detectors, achieving 75.87% mAP on the DOTA dataset and 96.50% mAP on the HRSC2016 dataset, while running at 15.1 FPS with the image size of 1024×1024 on a single RTX 2080Ti. Thus, the oriented R-CNN is a practical object detection framework in terms of both accuracy and efficiency. We hope our method will inspire rethinking the design of oriented object detectors and oriented object regression scheme, as well as serve as a solid baseline for oriented object detection. For future research, our code is available at https://github.com/jbwang1997/OBBDetection.

在没有附加功能的情况下，我们在两个公开数据集 DOTA、HRSC2016 上检测了 Oriented R-CNN 模型。使用 ResNet-50-FPN，在 DOTA 数据集上达到了 75.87% 的 mAP，在 HRSC2016 上达到了 96.50%的 mAP，在单个 RTX2080Ti 上处理 1024x1024 的图像达到了 15.1FPS 的帧率。因此，从准确性和效率上来看，Oriented R-CNN 是一个实用的目标检测框架。我们希望此方法能够启发大家重新思考有向目标检测器以及针对有向目标的回归方案的设计，并作为有向目标检测领域的一个 baseline。我们的代码开源在 https://github.com/jbwang1997/OBBDetection。